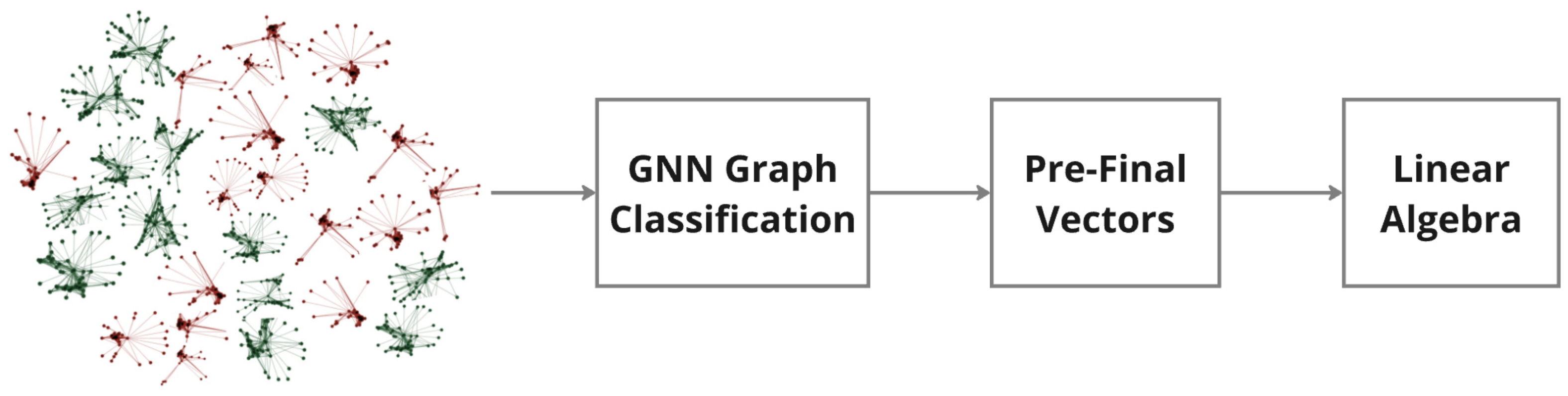
Utilizing Pre-Final Vectors from GNN Graph Classification for Enhanced Climate Analysis

ALEX ROMANOVA, Independent Researcher

Method: Small Graphs Embedding

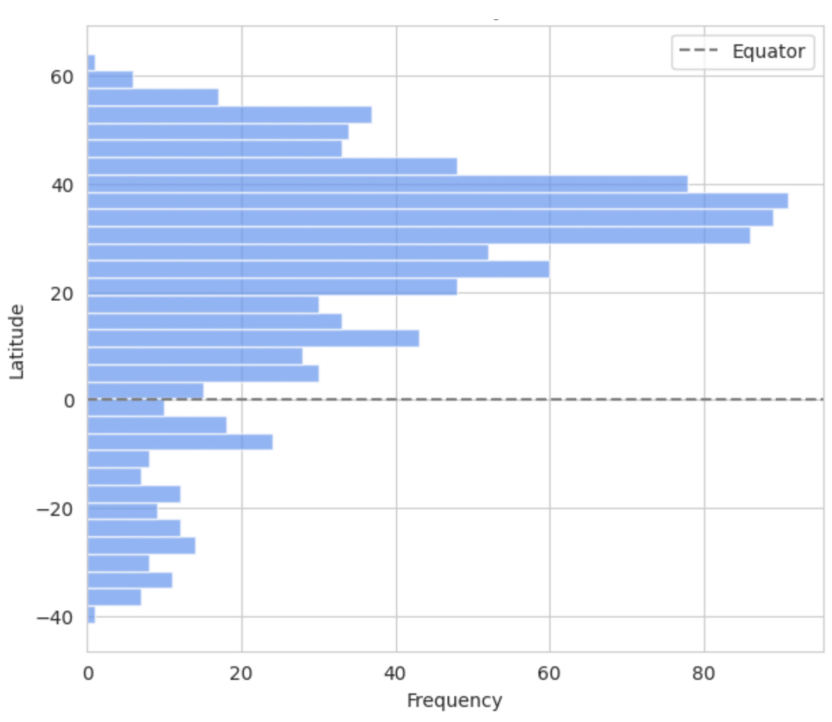
Model complex data as 'small graphs' and use GNN graph classification to convert them into vectors, enabling deeper analysis and unlocking actionable insights across diverse data domains.



Application Ideas

- **Social Networks:** Model persons of interest and their friends of friends as small graphs to analyze influencers and community structure.
- **Document Analysis:** Use semantic subgraphs to enhance topic modeling, text classification, and key term connections.
- **Sliding Window Graphs:** Apply to stock price time series to uncover anomalies like sudden price spikes and forecast emerging trends.

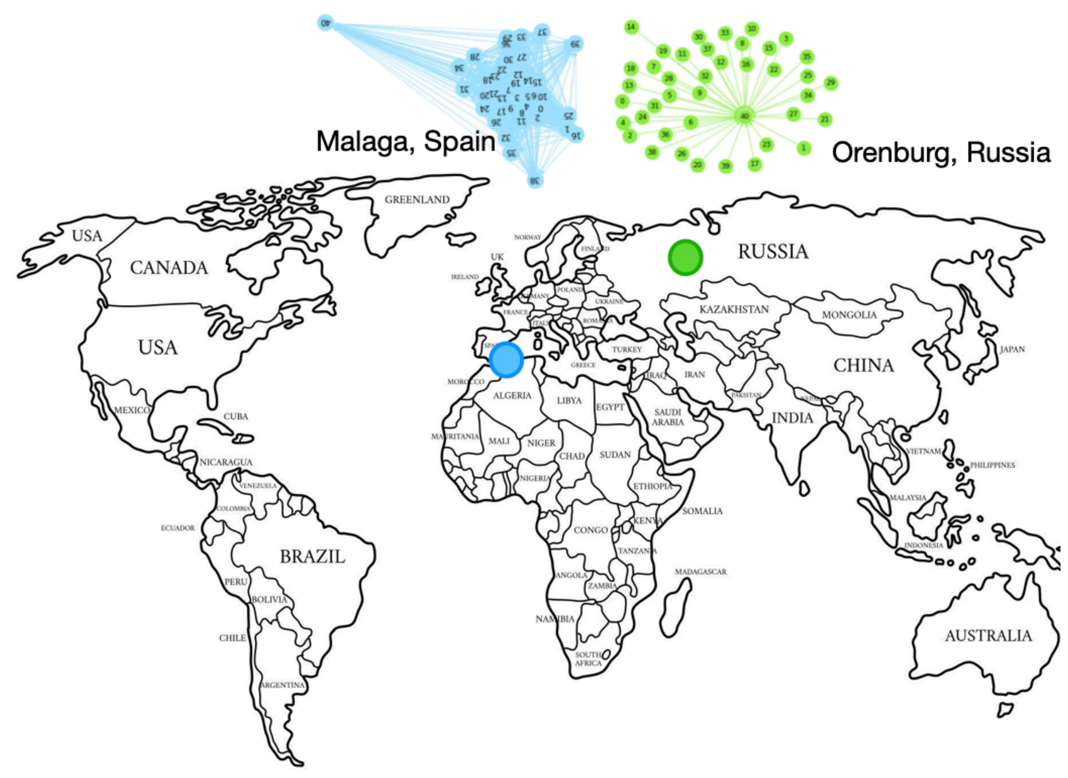
Climate Raw Data



- 40 years of daily temperatures for 1,000 major cities.
- Is the climate stable or unstable?

Data Preparation

- **Cosine Similarity Matrices:** Based on city-year climate data.
- **Graph Transformation:** Convert matrices into graphs using high cosine similarities.
- **Virtual Nodes:** Add to ensure each graph forms a single connected component.



Pre-Final Vectors

- **Graph Labeling:** Classify graphs as 'stable' or 'unstable' based on latitude.
- **Model Training:** Train GNN Graph Classification using PyTorch Geometric (PyG).
- **Pre-Final Vectors:** Extract vectors from the trained model.
- **Further Analysis:** Utilize embedded small graphs for deeper data insights.

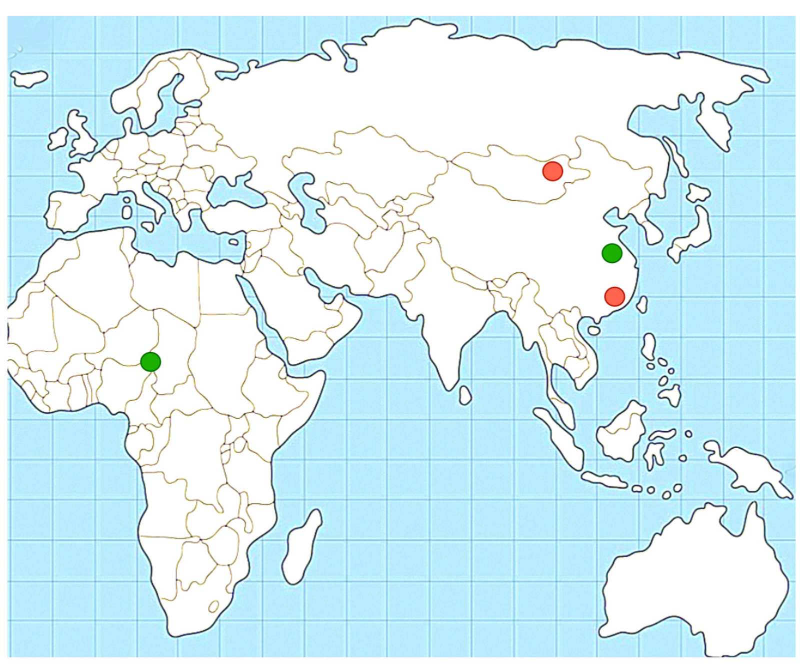
Cosine Similarity Analysis

The closest neighbors of city-graph vectors for Tokyo, Japan (largest city).

City	Country	Latitude	Longitude	Label
Shiraz	Iran	29.63	52.57	0
Beirut	Lebanon	33.87	35.51	1
Luoyang	China	34.68	112.47	1
Latakia	Syria	35.54	35.78	1
Sacramento	United States	38.57	-121.47	1

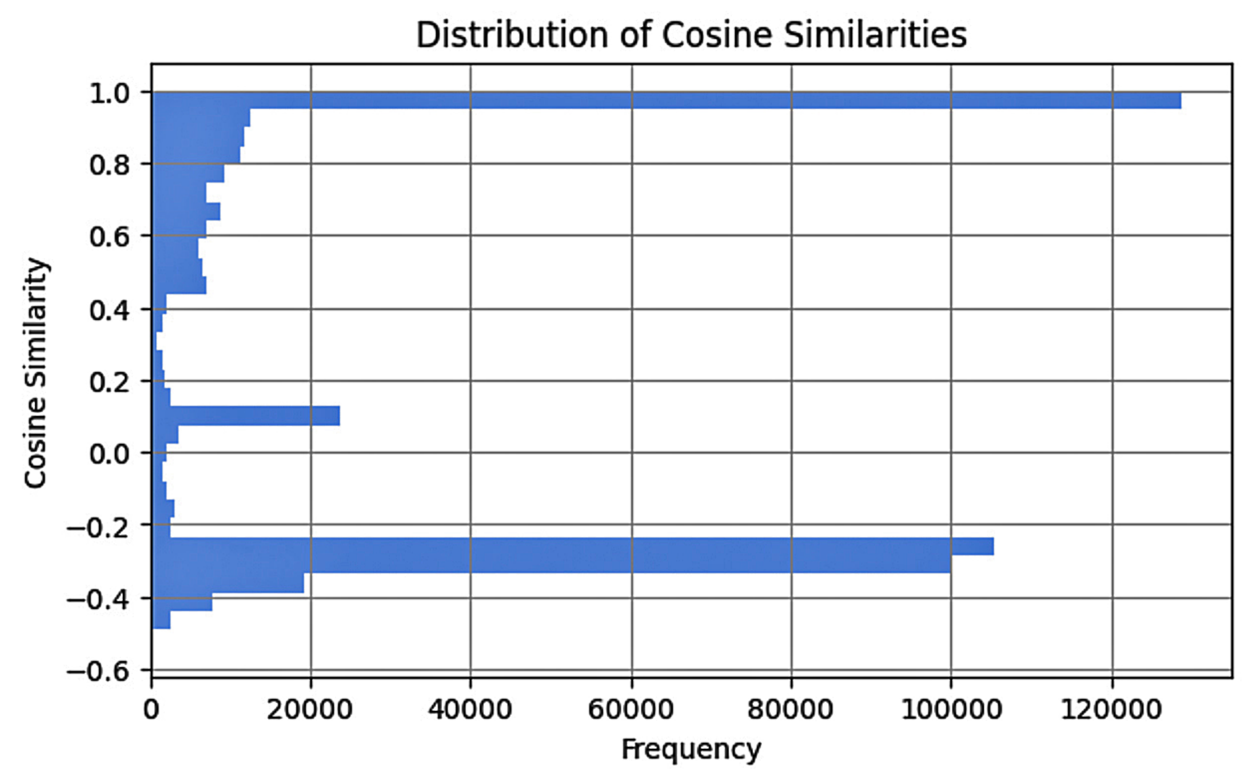
The closest neighbors of city-graph vectors for Gothenburg, Sweden (smallest city).

City	Country	Latitude	Longitude	Label
Krakow	Poland	50.06	19.96	1
Katowice	Poland	50.26	19.02	1
Bytom	Poland	50.35	18.91	1
Lodz	Poland	51.77	19.45	1
Gdansk	Poland	54.36	18.64	1



- **Lowest Cosine Similarity** (-0.543011): Ulaanbaatar, Mongolia, and Shaoguan, China.
- **Near-Orthogonal Cosine Similarity** (-0.000047): Nanchang, China, and N'Djamena, Chad.

Meta-Graphs Mining



- Set thresholds from cosine similarity distribution.
- **High Similarity Graph** (Threshold > 0.9): Shortest distance is 1.93 km (Jerusalem, Israel – Al Quds, West Bank).
- **Low Similarity Graph** (Threshold -0.4 to -0.2): Shortest distance is 115 km (Kabul – Jalalabad, Afghanistan).
- **Very High Similarity Graph** (Threshold > 0.99): connected components of sizes [514, 468, 7, 5]
- 514-node Component: Mostly stable climates
- 468-node Component: Mostly unstable climates.
- Small components show boundary cities between stable and unstable climates.